

# Lipschitz Continuity and Approximate Equilibria

Argyrios Deligkas<sup>1</sup>, John Fearnley<sup>1</sup>, and Paul Spirakis<sup>1,2</sup>

<sup>1</sup> University of Liverpool, UK

<sup>2</sup> Computer Technology Institute (CTI), Greece

**Abstract.** In this paper, we study games with continuous action spaces and non-linear payoff functions. Our key insight is that Lipschitz continuity of the payoff function allows us to provide algorithms for finding approximate equilibria in these games. We begin by studying Lipschitz games, which encompass, for example, all concave games with Lipschitz continuous payoff functions. We provide an efficient algorithm for computing approximate equilibria in these games. Then we turn our attention to penalty games, which encompass biased games and games in which players take risk into account. Here we show that if the penalty function is Lipschitz continuous, then we can provide a quasi-polynomial time approximation scheme. Finally, we study distance biased games, where we present simple strongly polynomial time algorithms for finding best responses in  $L_1$ ,  $L_2^2$ , and  $L_\infty$  biased games, and then use these algorithms to provide strongly polynomial algorithms that find  $2/3$ ,  $5/7$ , and  $2/3$  approximations for these norms, respectively.

## 1 Introduction

Nash equilibria [18] are the central solution concept in game theory. However, recent advances have shown that computing an *exact* Nash equilibrium is PPAD-complete [7, 8], and so there are unlikely to be polynomial time algorithms for this problem. The hardness of computing exact equilibria has lead to the study of *approximate* equilibria: while an exact equilibrium requires that all players have no incentive to deviate from their current strategy, an  $\epsilon$ -approximate equilibrium requires only that their incentive to deviate is less than  $\epsilon$ .

A fruitful line of work has developed studying the best approximations that can be found in polynomial-time for *bimatrix games*, which are two-player strategic form games. There, after a number of papers [4, 9, 10], the best known algorithm was given by Tsaknakis and Spirakis [20], who provide a polynomial time algorithm that finds a 0.3393-equilibrium. A prominent open problem is whether there exists a PTAS for this problem. The existence of an FPTAS was ruled out by Chen, Deng, and Teng [7] unless  $\text{PPAD} = \text{P}$ . While the existence of a PTAS remains open, there is however a *quasi-polynomial* approximation scheme given by Lipton, Markakis, and Mehta [16].

In a strategic form game, the game is specified by giving each player a finite number of strategies, and then specifying a table of payoffs that contains one entry for every possible combination of strategies that the players might pick.

The players are allowed to use mixed strategies, and so ultimately the payoff function is a convex combination of the payoffs given in the table. However, some games can only be modelled in a more general setting where the action spaces are continuous, or the payoff functions are non-linear.

For example, Rosen’s seminal work [19] considered *concave games*, where each player picks a vector from a convex set. The payoff to each player is specified by a function that satisfies the following condition: if every other player’s strategy is fixed, then the payoff to a player is a convex function over his strategy space. Rosen proved that concave games always have an equilibrium. A natural subclass of concave games, studied by Caragiannis, Kurokawa, and Procaccia [5], is the class of biased games. A biased game is defined by a strategic form game, a *base strategy* and a *penalty function*. The players play the strategic form game as normal, but they all suffer a penalty for deviating from their base strategy. This penalty can be a non-linear function, such as the  $L_2^2$  norm.

In this paper, we study the computation of approximate equilibria in such games. Our main observation is that Lipschitz continuity of the players’ payoff functions (with respect to changes in the strategy space) allows us to provide algorithms that find approximate equilibria. Several papers have studied how the Lipschitz continuity of the players’ payoff functions affects the existence, the quality, and the complexity of the equilibria of the underlying game. Azrieli and Shmaya [1] studied many player games and derived bounds for the Lipschitz constant of the utility functions for the players that guarantees the existence of pure approximate equilibria for the game. Daskalakis and Papadimitriou [11] proved that anonymous games possess pure approximate equilibria whose quality depends on the Lipschitz constant of the payoff functions and the number of pure strategies the players have and proved that these approximate equilibria can be computed in polynomial time. Furthermore, they gave a polynomial-time approximation scheme for anonymous games with many players and constant number of pure strategies. Babichenko [2] presented a best-reply dynamic for  $n$ -players Lipschitz anonymous games with two strategies which reaches an approximate pure equilibrium in  $O(n \log n)$  steps. Deb and Kalai [12] studied how some variants of the Lipschitz continuity of the utility functions are sufficient to guarantee hindsight stability of equilibria.

## 1.1 Our contribution

**Lipschitz games.** We begin by studying a very general class of games, where each player’s strategy space is continuous, and represented by a convex set of vectors, and where the only restriction is that the payoff function is Lipschitz continuous. This class encompasses, for example, every concave game in which the payoffs are Lipschitz continuous. This class is so general that exact equilibria, and even approximate equilibria may not exist. Nevertheless, we give an efficient algorithm that either outputs an  $\epsilon$ -equilibrium, or determines that the game has no exact equilibria. More precisely, for  $M$  player games with a strategy space defined as the convex hull of  $n$  vectors, that have  $\lambda$ -Lipschitz continuous payoff functions in the  $L_p$  norm, for  $p \geq 2$ , and where  $\gamma = \max \|\mathbf{x}\|_p$  over all

$\mathbf{x}$  in the strategy space, we either compute an  $\epsilon$ -equilibrium or determine that no exact equilibrium exists in time  $O(Mn^{Mk+l})$ , where  $k = O(\frac{\lambda^2 M p \gamma^2}{\epsilon^2})$  and  $l = O(\frac{\lambda^2 p \gamma^2}{\epsilon^2})$ . Observe that this is a polynomial time algorithm when  $\lambda$ ,  $p$ ,  $\gamma$ ,  $M$ , and  $\epsilon$  are constant.

To prove this result, we utilize a recent result of Barman [3], which states that for every vector in a convex set, there is another vector that is  $\epsilon$  close to the original in the  $L_p$  norm, and is a convex combination of  $b$  points on the convex hull, where  $b$  depends on  $p$  and  $\epsilon$ , but does not depend on the dimension. Using this result, and the Lipschitz continuity of the payoffs, allows us to reduce the task of finding an  $\epsilon$ -equilibrium to checking only a small number of strategy profiles, and thus we get a brute-force algorithm that is reminiscent of the QPTAS given by Lipton, Markakis, and Mehta for bimatrix games [16].

However, life is not so simple for us. Since we study a very general class of games, verifying whether a given strategy profile is an  $\epsilon$ -equilibrium is a non-trivial task. It requires us to compute a *regret* for each player, which is the difference between the player's best response payoff and their actual payoff. Computing a best response in a bimatrix game is trivial, but for Lipschitz games, it may be a hard problem. We get around this problem by instead giving an algorithm to compute *approximate* best responses. Hence we find *approximate* regrets, and it turns out that this is sufficient for our algorithm to work.

**Penalty games.** We then turn our attention to *penalty games*. In these games, the players play a strategic form game, and their utility is the payoff achieved in the game *minus* a penalty. The penalty function can be an arbitrary function that depends on the player's strategy. This is a general class of games that encompasses a number of games that have been studied before. The biased games studied by Caragiannis, Kurokawa, and Procaccia [5] are penalty games where the penalty is determined by the amount that a player deviates from a specified base strategy. The biased model was studied in the past by psychologists [21] and it is close to what they call *anchoring* [6, 14]. In their seminal paper, Fiat and Papadimitriou [13] introduced a model for *risk prone* games, which resemble penalty games since the risk component can be encoded as a penalty. Mavronicolas and Monien [17] followed this line of research and provided results on the complexity of deciding if such games possess an equilibrium.

We again show that Lipschitz continuity helps us to find approximate equilibria. The only assumption that we make is that the penalty function is Lipschitz continuous in an  $L_p$  norm with  $p \geq 2$ . Again, this is a weak restriction, and it does not guarantee that exact equilibria exist. Even so, we give a quasi-polynomial time algorithm that either finds an  $\epsilon$ -equilibrium, or verifies that the game has no exact equilibrium.

Our result can be seen as a generalisation of the QPTAS given by Lipton, Markakis, and Mehta [16] for bimatrix games. Their approach is to show the existence of an approximate equilibrium with a logarithmic support. They proved this via the probabilistic method: if we know an exact equilibrium of a bimatrix game, then we can take logarithmically many samples from the strategies, and playing the sampled strategies uniformly will be an approximate equilibrium with

positive probability. We take a similar approach, but since our games are more complicated, our proof is necessarily more involved. In particular, for Lipton, Markakis, and Mehta, proving that the sampled strategies are an approximate equilibrium only requires showing that the expected payoff is close to the best response payoff. In penalty games, best response strategies are not necessarily pure, and so the events that we must consider are more complex.

**Distance biased games.** Finally, we consider distance biased games, which form a subclass of penalty games that have been studied recently by Caragiannis, Kurokawa, and Procaccia [5]. They showed that, under very mild assumptions on the bias function, biased games always have an exact equilibrium. Furthermore, for the case where the bias function is either the  $L_1$  norm, or the  $L_2^2$  norm, they give an exponential time algorithm for finding an exact equilibrium.

Our results for penalty games already give a QPTAS for biased games, but we are also interested in whether there are polynomial-time algorithms that can find non-trivial approximations. We give a positive answer to this question for games where the bias is the  $L_1$  norm, the  $L_2^2$  norm, or the  $L_\infty$  norm. We follow the well-known approach of Daskalakis, Mehta, Papadimitriou [10], who gave a simple algorithm for finding a 0.5-approximate equilibrium in a bimatrix game.

We show that this algorithm also works for biased games, although the generalisation is not entirely trivial. Again, this is because best responses cannot be trivially computed in biased games. For the  $L_1$  and  $L_\infty$  norms, best responses can be computed via linear programming, and for the  $L_2^2$  norm, best responses can be formulated as a quadratic program, and it turns out that this particular QP can be solved in polynomial time by the ellipsoid method. However, none of these algorithms are strongly polynomial. We show that, for each of the norms, best responses can be found by a simple strongly-polynomial combinatorial algorithm. We then analyse the quality of approximation provided by the technique of Daskalakis, Mehta, Papadimitriou [10]. We obtain a strongly polynomial algorithm for finding a  $2/3$  approximation in  $L_1$  and  $L_\infty$  biased games, and a strongly polynomial algorithm for finding a  $5/7$  approximation in  $L_2^2$  biased games. For the latter result, in the special case where the bias function is the inner product of the player's strategy we find a  $13/21$  approximation.

## 2 Preliminaries

We start by fixing some notation. For each positive integer  $n$  we use  $[n]$  to denote the set  $\{1, 2, \dots, n\}$ , we use  $\Delta^n$  to denote the  $(n - 1)$ -dimensional simplex, and  $\|x\|_p$  to denote the  $p$ -norm of a vector  $x \in \mathbb{R}^d$ , i.e.  $\|x\|_p = (\sum_{i \in [d]} |x_i|^p)^{1/p}$ . Given a set  $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$ , we use  $\text{conv}(X)$  to denote the convex hull of  $X$ . A vector  $y \in \text{conv}(X)$  is said to be  $k$ -uniform with respect to  $X$  if there exists a size  $k$  multiset  $S$  of  $[n]$  such that  $y = \frac{1}{k} \sum_{i \in S} x_i$ . When  $X$  is clear from the context we will simply say that a vector is  $k$  uniform without mentioning that uniformity is with respect to  $X$ .

**Games and strategies.** A game with  $M$  players can be described by a set of available actions for each player and a utility function for each player that

depends both on his chosen action and the actions the rest of the players chose. For each player  $i \in [M]$  we use  $S_i$  to denote his set of available actions and we call it his *strategy space*. We will use  $x_i \in S_i$  to denote a specific action chosen by player  $i$  and we will call it as the *strategy* of player  $i$ . Furthermore, we use  $\mathbf{x} = (x_1, \dots, x_M)$  to denote a *strategy profile* of the game. We use  $T_i(x_i, \mathbf{x}_{-i})$  to denote the utility of player  $i$  when he plays the strategy  $x_i$  and the rest of the players play according to the strategy profile  $\mathbf{x}_{-i}$ . A strategy  $\hat{x}_i$  is a *best response* against the strategy profile  $\mathbf{x}_{-i}$ , if  $T_i(\hat{x}_i, \mathbf{x}_{-i}) \geq T_i(x_i, \mathbf{x}_{-i})$  for all  $x_i \in S_i$ . The *regret* player  $i$  suffers under a strategy profile  $\mathbf{x}$  is the difference between the utility of his best response and his utility under  $\mathbf{x}$ , i.e.  $T_i(\hat{x}_i, \mathbf{x}_{-i}) - T_i(x_i, \mathbf{x}_{-i})$ .

**$\lambda_p$ -Lipschitz Games.** We will use the notion of the  $\lambda_p$ -Lipschitz continuity.

**Definition 1 ( $\lambda_p$ -Lipschitz).** A function  $f : A \rightarrow \mathbb{R}$ , with  $A \subseteq \mathbb{R}^d$  is  $\lambda_p$ -Lipschitz continuous if for every  $x$  and  $y$  in  $A$ , it is true that  $|f(x) - f(y)| \leq \lambda \cdot \|x - y\|_p$ .

We call the game  $\mathfrak{L} := (M, n, \lambda, p, \gamma, \mathcal{T})$   $\lambda_p$ -Lipschitz if for each player  $i \in [M]$  the strategy space  $S_i$  is the convex hull of  $n$  vectors  $y_1, \dots, y_n$  in  $\mathbb{R}^d$ ,  $\max_{x_i \in S_i} \|x_i\|_p \leq \gamma$ , and the utility function  $T_i(\mathbf{x}) \in \mathcal{T}$  is  $\lambda_p$ -Lipschitz continuous.

**Two Player Penalty Games.** A two player penalty game  $\mathcal{P}$  is defined by a tuple  $(R, C, \mathbf{f}_r(\mathbf{x}), \mathbf{f}_c(\mathbf{y}))$ , where  $(R, C)$  is a bimatrix game and  $\mathbf{f}_r(\mathbf{x})$  and  $\mathbf{f}_c(\mathbf{y})$  are the penalty functions for the row and the column player respectively. The utilities for the players under a strategy profile  $(\mathbf{x}, \mathbf{y})$ , denoted by  $T_r(\mathbf{x}, \mathbf{y})$  and  $T_c(\mathbf{x}, \mathbf{y})$ , are given by  $T_r(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T R \mathbf{y} - \mathbf{f}_r(\mathbf{x})$  and  $T_c(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T C \mathbf{y} - \mathbf{f}_c(\mathbf{y})$ . We will use  $\mathcal{P}_\lambda$  to denote the set of two player penalty games with  $\lambda_p$ -Lipschitz penalty functions. A special class of penalty games is obtained when  $\mathbf{f}_r(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$  and  $\mathbf{f}_c(\mathbf{y}) = \mathbf{y}^T \mathbf{y}$ . We call these games as *inner product* penalty games.

**Two Player Biased Games.** This is a subclass of penalty games, where extra constraints are added to the penalty functions  $\mathbf{f}_r(\mathbf{x})$  and  $\mathbf{f}_c(\mathbf{y})$  of the players. In this class of games there is a *base strategy* and for each player and the penalty they receive is increasing with the distance between the strategy they choose and their base strategy. Formally, the row player has a base strategy  $\mathbf{p} \in \Delta^n$ , the column player has a base strategy  $\mathbf{q}$  and their strictly increasing penalty functions are defined as  $\mathbf{f}_r(\|\mathbf{x} - \mathbf{p}\|_i^s)$  and  $\mathbf{f}_c(\|\mathbf{y} - \mathbf{q}\|_m^l)$  respectively.

**Two Player Distance Biased Games.** This is a special class of biased games where the penalty function is a fraction of the distance between the base strategy of the player and his chosen strategy. Formally, a two player distance biased game  $\mathcal{B}$  is defined by a tuple  $(R, C, \mathbf{b}_r(\mathbf{x}, \mathbf{p}), \mathbf{b}_c(\mathbf{y}, \mathbf{q}), d_r, d_c)$ , where  $(R, C)$  is a bimatrix game,  $\mathbf{p} \in \Delta^n$  is a base strategy for the row player,  $\mathbf{q} \in \Delta^n$  is a base strategy for the column player,  $\mathbf{b}_r(\mathbf{x}, \mathbf{p}) = \|\mathbf{x} - \mathbf{p}\|_i^s$  and  $\mathbf{b}_c(\mathbf{y}, \mathbf{q}) = \|\mathbf{y} - \mathbf{q}\|_m^l$  are the penalty functions for the row and the column player respectively. The utilities for the players under a strategy profile  $(\mathbf{x}, \mathbf{y})$ , denoted by  $T_r(\mathbf{x}, \mathbf{y})$  and  $T_c(\mathbf{x}, \mathbf{y})$ , are given by  $T_r(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T R \mathbf{y} - d_r \cdot \mathbf{b}_r(\mathbf{x}, \mathbf{p})$  and  $T_c(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T C \mathbf{y} - d_c \cdot \mathbf{b}_c(\mathbf{y}, \mathbf{q})$ , where  $d_r$  and  $d_c$  are non negative constants.

**Solution Concepts.** A strategy profile is an equilibrium if no player can increase his utility by unilaterally changing his strategy. A relaxed version of this

concept is the approximate equilibrium, or  $\epsilon$ -equilibrium, in which no player can increase his utility more than  $\epsilon$  by unilaterally changing his strategy. Formally, a strategy profile  $\mathbf{x}$  is an  $\epsilon$ -equilibrium in a game  $\mathfrak{L}$  if for every player  $i \in [M]$  it holds that  $T_i(x_i, \mathbf{x}_{-i}) \geq T_i(x'_i, \mathbf{x}_{-i}) - \epsilon$  for all  $x'_i \in S_i$ .

In [7] it was proven that, unless  $\mathsf{P} = \mathsf{PPAD}$ , there is no FPTAS for computing an  $\epsilon$ -NE in bimatrix games. The same result holds for the class of penalty games where the penalty functions  $f$  for the players depend on  $n$ , the size of the underlying bimatrix game, and  $\lim_{n \rightarrow \infty} f = 0$  for every player. Let  $\mathcal{P}'$  to denote this class of games.

**Theorem 1.** *Unless  $\mathsf{P} = \mathsf{PPAD}$ , there is no FPTAS for computing an  $\epsilon$ -equilibrium in penalty games in  $\mathcal{P}'$ .*

### 3 Approximate equilibria in $\lambda_p$ -Lipschitz games

In this section, we give an algorithm for computing approximate equilibria in  $\lambda_p$  Lipschitz games. Note that, our definition of a  $\lambda_p$ -Lipschitz game does not guarantee that an equilibrium always exists. Our technique can be applied *irrespective* of whether an exact equilibrium exists. If an exact equilibrium does exist, then our technique will always find an  $\epsilon$ -equilibrium. If an exact equilibrium does not exist, then our algorithm either finds an  $\epsilon$ -equilibrium or reports that the game does not have an exact equilibrium.

We will utilize the following theorem that was recently proved by Barman [3].

**Theorem 2 ([3]).** *Given a set of vectors  $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$ , let  $\text{conv}(X)$  denote the convex hull of  $X$ . Furthermore, let  $\gamma := \max_{x \in X} \|x\|_p$  for some  $2 \leq p < \infty$ . For every  $\epsilon > 0$  and every  $\mu \in \text{conv}(X)$ , there exists an  $\frac{4p\gamma^2}{\epsilon^2}$  uniform vector  $\mu' \in \text{conv}(X)$  such that  $\|\mu - \mu'\|_p \leq \epsilon$ .*

Combining Theorem 2 with the Definition 1 we get the following lemma.

**Lemma 1.** *Let  $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$ , let  $f : \text{conv}(X) \rightarrow \mathbb{R}$  be a  $\lambda_p$ -Lipschitz continuous function for some  $2 \leq p < \infty$ , let  $\epsilon > 0$  and let  $k = \frac{4\lambda^2 p \gamma^2}{\epsilon^2}$ , where  $\gamma := \max_{x \in X} \|x\|_p$ . Furthermore, let  $f(\mathbf{x}^*)$  be the optimum value of  $f$ . Then we can compute a  $k$ -uniform point  $\mathbf{x}' \in \text{conv}(X)$  in time  $O(n^k)$ , such that  $|f(\mathbf{x}^*) - f(\mathbf{x}')| < \epsilon$ .*

We now prove our result about Lipschitz games. In what follows we will study a  $\lambda_p$ -Lipschitz game  $\mathfrak{L} := (M, n, \lambda, p, \gamma, \mathcal{T})$ . Assuming the existence of an exact Nash equilibrium, we establish the existence of a  $k$ -uniform approximate equilibrium in the game  $\mathfrak{L}$ , where  $k$  depends on  $M, \lambda, p$  and  $\gamma$ . Note that  $\lambda$  depends heavily on  $p$  and the utility functions for the players.

Since by the definition of  $\lambda_p$ -Lipschitz games the strategy space  $S_i$  for every player  $i$  is the convex hull of  $n$  vectors  $y_1, \dots, y_n$  in  $\mathbb{R}^d$ , any  $x_i \in S_i$  can be written as a convex combination of  $y_j$ s. Hence,  $x_i = \sum_{j=1}^n \alpha_j y_j$ , where  $\alpha_j > 0$  for every  $j \in [n]$  and  $\sum_{j=1}^n \alpha_j = 1$ . Then,  $\alpha = (\alpha_1, \dots, \alpha_n)$  is a probability

distribution over the vectors  $y_1, \dots, y_n$ , i.e. vector  $y_j$  is drawn with probability  $\alpha_j$ . Thus, we can sample a strategy  $x_i$  by the probability distribution  $\alpha$ .

So, let  $\mathbf{x}^*$  be an equilibrium for  $\mathfrak{L}$  and let  $\mathbf{x}'$  be a sampled uniform strategy profile from  $\mathbf{x}^*$ . For each player  $i$  we define the following events

$$\begin{aligned}\phi_i &= \{|T_i(x'_i, \mathbf{x}'_{-i}) - T_i(x_i^*, \mathbf{x}_{-i}^*)| < \epsilon/2\} \\ \pi_i &= \{T_i(x_i, \mathbf{x}'_{-i}) < T_i(x'_i, \mathbf{x}'_{-i}) + \epsilon\} \quad \text{for all possible } x_i \\ \psi_i &= \{\|x'_i - x_i^*\|_p < \frac{\epsilon}{2M\lambda}\} \quad \text{for some } p > 0.\end{aligned}$$

Notice that if all the events  $\pi_i$  occur at the same time, then the sampled profile  $\mathbf{x}'$  is an  $\epsilon$ -equilibrium. We will show that if for a player  $i$  the events  $\phi_i$  and  $\bigcap_j \psi_j$  hold, then the event  $\pi_i$  is also true.

**Lemma 2.** *For all  $i \in [M]$  it holds that  $\bigcap_{j \in [M]} \psi_j \cap \phi_i \subseteq \pi_i$ .*

We are ready to prove the main result of the section.

**Theorem 3.** *In any game  $\lambda_p$ -Lipschitz game  $\mathfrak{L}$  that posses an equilibrium and any  $\epsilon > 0$ , there is a  $k$ -uniform strategy profile, with  $k = \frac{16M^2\lambda^2p\gamma^2}{\epsilon^2}$  that is an  $\epsilon$ -equilibrium.*

Theorem 3 establishes the existence of a  $k$ -uniform approximate equilibrium, but this does not immediately give us our approximation algorithm. The obvious approach is to perform a brute force check of all  $k$ -uniform strategies, and then output the one the provides the best approximation. There is a problem with this, however, since computing the quality of approximation requires us to compute the regret for each player, which in turn requires us to compute a best response for each player. Computing an exact best response in a Lipschitz game is a hard problem in general, since we make no assumptions about the utility functions of the players. Fortunately, it is sufficient to instead compute an *approximate* best response for each player, and Lemma 1 can be used to do this.

**Lemma 3.** *Let  $\mathbf{x}$  be a strategy profile for a  $\lambda_p$ -Lipschitz game  $\mathfrak{L}$ , and let  $\hat{x}_i$  be a best response for player  $i$  against the profile  $\mathbf{x}_{-i}$ . There is a  $\frac{4\lambda^2p\gamma^2}{\epsilon^2}$ -uniform strategy  $x'_i$  that is an  $\epsilon$ -best response against  $\mathbf{x}_{-i}$ .*

Our goal is to *approximate* the approximation guarantee for a given strategy profile. More formally, given a strategy profile  $\mathbf{x}$  that is an  $\epsilon$ -equilibrium, and a constant  $\delta > 0$ , we want an algorithm that outputs a number within the range  $[\epsilon - \delta, \epsilon + \delta]$ . Lemma 3 allows us to do this. For a given strategy profile  $\mathbf{x}$ , we first compute  $\delta$ -approximate best responses for each player, then we can use these to compute  $\delta$ -approximate regrets for each player. The maximum over the  $\delta$ -approximate regrets then gives us an approximation of  $\epsilon$  with a tolerance of  $\delta$ . This is formalised in the following algorithm.

**Algorithm 1.** Evaluation of approximation guarantee

**Input:** A strategy profile  $\mathbf{x}$  for  $\mathfrak{L}$ , and a constant  $\delta > 0$ .

**Output:** An additive  $\delta$ -approximation of the approximation guarantee  $\alpha(\mathbf{x})$  for the strategy profile  $\mathbf{x}$ .

1. Set  $l = \frac{4\lambda^2 p \gamma^2}{\delta^2}$ .
2. For every player  $i \in [M]$ 
  - (a) For every  $l$ -uniform strategy  $x'_i$  of player  $i$  compute  $T_i(x'_i, \mathbf{x}_{-i})$ .
  - (b) Set  $m^* = \max_{x'_i} T_i(x'_i, \mathbf{x}_{-i})$ .
  - (c) Set  $\mathcal{R}_i(\mathbf{x}) = m^* - T_i(x_i, \mathbf{x}_{-i})$ .
3. Set  $\alpha(\mathbf{x}) = \delta + \max_{i \in [M]} \mathcal{R}_i(\mathbf{x})$ .
4. Return  $\alpha(\mathbf{x})$ .

Utilising the above algorithm, we can now produce an algorithm to find an approximate equilibrium in Lipschitz games. The algorithm checks all  $k$ -uniform strategy profiles, using the value of  $k$  given by Theorem 3, and for each one, computes an approximation of the quality approximation using the algorithm given above.

**Algorithm 2.**  $3\epsilon$ -equilibrium for  $\lambda_p$ -Lipschitz game  $\mathfrak{L}$

**Input:** Game  $\mathfrak{L}$  and  $\epsilon > 0$ .

**Output:** An  $3\epsilon$ -equilibrium for  $\mathfrak{L}$ .

1. Set  $k > \frac{16\lambda^2 M p \gamma^2}{\epsilon^2}$ .
2. For every  $k$ -uniform strategy profile  $\mathbf{x}'$ 
  - (a) Compute an  $\epsilon$ -approximation of  $\alpha(\mathbf{x}')$ .
  - (b) If the  $\epsilon$ -approximation of  $\alpha(\mathbf{x}')$  is less than  $2\epsilon$ , return  $\mathbf{x}'$ .

If the algorithm returns a strategy profile  $\mathbf{x}$ , then it must be a  $3\epsilon$  equilibrium. This is because we check that an  $\epsilon$ -approximation of  $\alpha(\mathbf{x})$  is less than  $2\epsilon$ , and therefore  $\alpha(\mathbf{x}) \leq 3\epsilon$ . Secondly, we argue that if the game has an exact Nash equilibrium, then this procedure will always output a  $3\epsilon$ -approximate equilibrium. From Theorem 3 we know that if  $k > \frac{16\lambda^2 M p \gamma^2}{\epsilon^2}$ , then there is a  $k$ -uniform strategy profile  $\mathbf{x}$  that is an  $\epsilon$ -equilibrium for  $\mathfrak{L}$ . When we apply our approximate regret algorithm to  $\mathbf{x}$ , to find an  $\epsilon$ -approximation of  $\alpha(\mathbf{x})$ , the algorithm will return a number that is less than  $2\epsilon$ , hence  $\mathbf{x}$  will be returned by the algorithm.

To analyse the running time, observe that there are  $\binom{n+k-1}{k} = O(n^k)$  possible  $k$ -uniform strategies for each player, thus  $O(n^{Mk})$   $k$ -uniform strategy profiles. Furthermore, our regret approximation algorithm runs in time  $O(Mn^l)$ , where  $l = \frac{4\lambda^2 p \gamma^2}{\epsilon^2}$ . Hence, we get the next theorem.

**Theorem 4.** *Given a  $\lambda_p$ -Lipschitz game  $\mathfrak{L}$  that posses an equilibrium and any  $\epsilon > 0$ , a  $3\epsilon$ -equilibrium can be computed in time  $O(Mn^{Mk+l})$ , where  $k = O(\frac{\lambda^2 M p \gamma^2}{\epsilon^2})$  and  $l = O(\frac{\lambda^2 p \gamma^2}{\epsilon^2})$ .*



Although it might be hard to decide whether a game has an equilibrium, our algorithm can be applied in *any*  $\lambda_p$ -Lipschitz game. Notice that our algorithm never uses the fact that the game possesses an equilibrium. If the game does not possess an exact equilibrium then our algorithm either finds an approximate equilibrium or determines that the game does not possess an exact equilibrium.

**Theorem 5.** *For any game  $\lambda_p$ -Lipschitz game  $\mathfrak{L}$  in time  $O(Mn^{Mk+l})$ , we can either compute a  $3\epsilon$ -equilibrium, or decide that  $\mathfrak{L}$  does not possess an exact equilibrium, where  $k = O(\frac{\lambda^2 Mp\gamma^2}{\epsilon^2})$  and  $l = O(\frac{\lambda^2 p\gamma^2}{\epsilon^2})$ .*

## 4 A quasi-polynomial algorithm for penalty games

In this section we present an algorithm that, for any  $\epsilon > 0$ , can compute an  $\epsilon$ -equilibrium for any penalty game in  $\mathcal{P}_\lambda$  that possesses one in quasi-polynomial time. For the algorithm, we take the same approach as we did in the previous section for Lipschitz games: we show that if an exact equilibrium exists, then a  $k$ -uniform approximate equilibrium always exists too, and provide a brute-force search algorithm for finding it. Once again, since best response computation may be hard for this class of games, we must provide an approximation algorithm for finding the quality of an approximate equilibrium.

We first focus on penalty games that possess an exact equilibrium. So, let  $(\mathbf{x}^*, \mathbf{y}^*)$  be an equilibrium of the game and let  $(\mathbf{x}', \mathbf{y}')$  be a  $k$ -uniform strategy profile sampled from this equilibrium. We define the following four events:

$$\begin{aligned}\phi_r &= \{|T_r(\mathbf{x}', \mathbf{y}') - T_r(\mathbf{x}^*, \mathbf{y}^*)| < \epsilon/2\} \\ \pi_r &= \{T_r(\mathbf{x}, \mathbf{y}') < T_r(\mathbf{x}', \mathbf{y}') + \epsilon\} \quad \text{for all } \mathbf{x} \\ \phi_c &= \{|T_c(\mathbf{x}', \mathbf{y}') - T_c(\mathbf{x}^*, \mathbf{y}^*)| < \epsilon/2\} \\ \pi_c &= \{T_c(\mathbf{x}', \mathbf{y}) < T_c(\mathbf{x}', \mathbf{y}') + \epsilon\} \quad \text{for all } \mathbf{y}.\end{aligned}$$

The goal is to derive a value for  $k$  such that all the four events above are true, or equivalently  $Pr(\phi_r \cap \pi_r \cap \phi_c \cap \pi_c) > 0$ .

Note that in order to prove that  $(\mathbf{x}', \mathbf{y}')$  is an  $\epsilon$ -equilibrium we *only* have to consider the events  $\pi_r$  and  $\pi_c$ . Nevertheless, as we show in Lemma 4, the events  $\phi_r$  and  $\phi_c$  are crucial in our analysis. The proof of the main theorem boils down to the events  $\phi_r$  and  $\phi_c$ .

We will focus only on the row player, since the same analysis can be applied to the column player. Firstly we study the event  $\pi_r$ .

**Lemma 4.** *For all penalty games it holds that  $Pr(\pi_r^c) \leq n \cdot e^{-\frac{k\epsilon^2}{2}} + Pr(\phi_r^c)$ .*

With Lemma 4 in hand, we can see that in order to compute a value for  $k$  it is sufficient to study the event  $\phi_r$ . We introduce the following auxiliary events that we will study separately:  $\phi_{ru} = \{|\mathbf{x}'^T R \mathbf{y}' - \mathbf{x}^{*T} R \mathbf{y}^*| < \epsilon/4\}$  and  $\phi_{rb} = \{|\mathbf{f}_r(\mathbf{x}') - \mathbf{f}_r(\mathbf{x}^*)| < \epsilon/4\}$ . It is easy to see that if both  $\phi_{rb}$  and  $\phi_{ru}$  are true, then the event  $\phi_r$  must be true too. So we have  $\phi_{rb} \cap \phi_{ru} \subseteq \phi_r$ . Using the analysis from [16] we can prove that  $Pr(\phi_{ru}^c) \leq 2e^{-\frac{k\epsilon^2}{8}}$ . Finally, we must prove an upper bound on the event  $\phi_{rb}^c$ , which we provide in the following lemma.

**Lemma 5.**  $\Pr(\phi_{rb}^c) \leq \frac{8\lambda\sqrt{p}}{\epsilon\sqrt{k}}.$

Let us define the event  $GOOD = \phi_r \cap \phi_c \cap \pi_r \cap \pi_c$ . To prove our theorem it suffices to prove that  $\Pr(GOOD) > 0$ . Notice that for the events  $\phi_c$  and  $\pi_c$  the same analysis as for  $\phi_r$  and  $\pi_r$  can be used. Then, using Lemma 4, Lemma 5 and the analysis for  $\phi_{ru}$  we get that  $\Pr(GOOD^c) < 1$  for the chosen value of  $k$ .

**Theorem 6.** *For any equilibrium  $(\mathbf{x}^*, \mathbf{y}^*)$  of a penalty game from the class  $\mathcal{P}_\lambda$ , any  $\epsilon > 0$ , and any  $k \in \frac{\Omega(\lambda^2 \log n)}{\epsilon^2}$ , there exists a  $k$ -uniform strategy profile  $(\mathbf{x}', \mathbf{y}')$  that:*

1.  $(\mathbf{x}', \mathbf{y}')$  is an  $\epsilon$ -equilibrium for the game,
2.  $|T_r(\mathbf{x}', \mathbf{y}') - T_r(\mathbf{x}^*, \mathbf{y}^*)| < \epsilon/2$ ,
3.  $|T_c(\mathbf{x}', \mathbf{y}') - T_c(\mathbf{x}^*, \mathbf{y}^*)| < \epsilon/2$ .

Theorem 6 establishes the *existence* of a  $k$ -uniform strategy profile  $(\mathbf{x}', \mathbf{y}')$  that is an  $\epsilon$ -equilibrium, but as before, we must provide an efficient method for approximating the quality of approximation provided by a given strategy profile. To do so, we first give the following lemma, which shows that approximate best responses can be computed in quasi-polynomial time for penalty games.

**Lemma 6.** *Let  $(\mathbf{x}, \mathbf{y})$  be a strategy profile for a penalty game  $\mathcal{P}_\lambda$ , and let  $\hat{\mathbf{x}}$  be a best response against  $\mathbf{y}$ . There is an  $l$ -uniform strategy  $\mathbf{x}'$ , with  $l = \frac{17\lambda^2\sqrt{p}}{\epsilon^2}$ , that is an  $\epsilon$ -best response against  $\mathbf{y}$ , i.e.  $T_r(\hat{\mathbf{x}}, \mathbf{y}) < T_r(\mathbf{x}', \mathbf{y}) + \epsilon$ .*

Having given this lemma, we can reuse Algorithm 1, but with  $l$  set equal to  $\frac{17\lambda^2\sqrt{p}}{\epsilon^2}$ , to provide an algorithm that approximates the quality of approximation of a given strategy profile. Then, we can reuse Algorithm 2 with  $k = \frac{\Omega(\lambda^2 \log n)}{\epsilon^2}$  to provide a quasi-polynomial time algorithm that finds approximate equilibria in penalty games. Notice again that our algorithm can be applied in games in which it is computationally hard to verify whether an exact equilibrium exists. Our algorithm either will compute an approximate equilibrium or it will fail to find one, in which case the game does not possess an exact equilibrium.

**Theorem 7.** *In any penalty game  $\mathcal{P}_\lambda$  and any  $\epsilon > 0$ , in quasi polynomial time we can either compute a  $3\epsilon$ -equilibrium, or decide that  $\mathcal{P}_\lambda$  does not possess an exact equilibrium.*

## 5 Distance Biased Games

In this section, we focus on three particular classes of distance biased games, and we provide polynomial-time approximation algorithms when the penalty function is one of the  $L_1, L_2^2$  and  $L_\infty$  norm. Our approach is to follow the technique of Daskalakis, Mehta, Papadimitriou [10] that finds a 0.5-NE in a bimatrix game. The algorithm that we will use for all three penalty functions is given below.

#### Algorithm 4. The Base Algorithm

1. Compute a best response  $\mathbf{y}^*$  against  $\mathbf{p}$ .
2. Compute a best response  $\mathbf{x}$  against  $\mathbf{y}^*$ .
3. Set  $\mathbf{x}^* = \delta \cdot \mathbf{p} + (1 - \delta) \cdot \mathbf{x}$ , for some  $\delta \in [0, 1]$ .
4. Return the strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$ .

While this is a well-known technique for bimatrix games, it cannot immediately be applied to penalty games, because the algorithm requires us to compute two best responses. While computing a best-response is trivial in bimatrix games, this is not the case for penalty games. Best responses for  $L_1$  and  $L_\infty$  penalties can be computed in polynomial-time via linear programming, and for  $L_2^2$  penalties, the ellipsoid algorithm can be applied to a specialized quadratic program. However, these methods work as black boxes and do not provide strongly polynomial algorithms.

For each of the penalties we develop a simple combinatorial algorithm for computing best response strategies. We use the nature of these penalty functions and we provide strongly polynomial algorithms that compute best responses. More specifically, for the  $L_1$  and  $L_\infty$  norms we compute the exact probability each pure strategy should be played in a best response by studying how the utility function increases. For the  $L_2^2$  norm we use the KKT conditions of a quadratic program to produce a closed formula for the solution. Our algorithms, which are strongly polynomial, allow us to optimize the value of  $\delta$ , and produce the following approximation guarantees.

**Theorem 8.** *In biased games with  $L_1, L_2^2$  and  $L_\infty$  penalties a  $2/3, 5/7$  and  $2/3$ -equilibrium respectively can be computed in polynomial time. For inner product games the approximation guarantee is  $13/21$ .*

## 6 Conclusions

We have studied games with infinite action spaces, and non-linear payoff functions. We have shown that Lipschitz continuity of the payoff function can be exploited to provide approximation algorithms. For Lipschitz games, Lipschitz continuity of the payoff function allows us to provide an efficient algorithm for finding approximate equilibria. For penalty games, Lipschitz continuity of the penalty function allows us to provide a QPTAS. Finally, we provided strongly polynomial approximation algorithms for  $L_1, L_2^2$ , and  $L_\infty$  distance biased games.

## References

1. Y. Azrieli and E. Shmaya. Lipschitz games. *Math. Oper. Res.*, 38(2):350–357, 2013.
2. Y. Babichenko. Best-reply dynamics in large binary-choice anonymous games. *Games and Economic Behavior*, 81:130–144, 2013.

3. S. Barman. Approximating Nash equilibria and dense bipartite subgraphs via an approximate version of Caratheodory's theorem. In *Proc. of STOC 2015*, pages 361–369, 2015.
4. H. Bosse, J. Byrka, and E. Markakis. New algorithms for approximate Nash equilibria in bimatrix games. *Theoretical Computer Science*, 411(1):164–173, 2010.
5. I. Caragiannis, D. Kurokawa, and A. D. Procaccia. Biased games. In *Proc. of AAAI*, pages 609–615, 2014.
6. G. B. Chapman and E. J. Johnson. Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, 79(2):115 – 153, 1999.
7. X. Chen, X. Deng, and S.-H. Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3):14:1–14:57, 2009.
8. C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.
9. C. Daskalakis, A. Mehta, and C. H. Papadimitriou. Progress in approximate Nash equilibria. In *Proc. of EC*, pages 355–358, 2007.
10. C. Daskalakis, A. Mehta, and C. H. Papadimitriou. A note on approximate Nash equilibria. *Theoretical Computer Science*, 410(17):1581–1588, 2009.
11. C. Daskalakis and C. H. Papadimitriou. Approximate Nash equilibria in anonymous games. *Journal of Economic Theory*, 2014. To appear.
12. J. Deb and E. Kalai. Stability in large Bayesian games with heterogeneous players. *Journal of Economic Theory*, 157(C):1041–1055, 2015.
13. A. Fiat and C. H. Papadimitriou. When the players are not expectation maximizers. In *SAGT*, pages 1–14, 2010.
14. D. Kahneman. Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, 51(2):296–312, 1992.
15. M. Kozlov, S. Tarasov, and L. Khachiyan. The polynomial solvability of convex quadratic programming. *{USSR} Computational Mathematics and Mathematical Physics*, 20(5):223 – 228, 1980.
16. R. J. Lipton, E. Markakis, and A. Mehta. Playing large games using simple strategies. In *EC*, pages 36–41, 2003.
17. M. Mavronicolas and B. Monien. The complexity of equilibria for risk-modeling valuations. *CoRR*, abs/1510.08980, 2015.
18. J. Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951.
19. J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):pp. 520–534, 1965.
20. H. Tsaknakis and P. G. Spirakis. An optimization approach for approximate Nash equilibria. *Internet Mathematics*, 5(4):365–382, 2008.
21. A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

## A Proof of Theorem 1

*Proof.* For the sake of contradiction suppose that there is an FPTAS for computing an  $\epsilon$ -equilibrium for penalty games in  $\mathcal{P}'$ . Then given an  $n \times n$  bimatrix game  $(R, C)$ , define the penalty game  $(R, C, f_r(\mathbf{x}), f_c(\mathbf{y}))$  from the family  $\mathcal{P}'$  where  $\lim_{n \rightarrow \infty} f_r(\mathbf{x}) = 0$  and  $\lim_{n \rightarrow \infty} f_c(\mathbf{y}) = 0$ . Let  $(\mathbf{x}^*, \mathbf{y}^*)$  be an  $\epsilon$ -equilibrium for the penalty game. This means that for all  $\mathbf{x}' \in \Delta^n$  it holds that  $\mathbf{x}^{*T} R \mathbf{y}^* - f_r(\mathbf{x}^*) \geq \mathbf{x}'^T R \mathbf{y}^* - f_r(\mathbf{x}') - \epsilon$  or, equivalently,  $\mathbf{x}^{*T} R \mathbf{y}^* \geq \mathbf{x}'^T R \mathbf{y}^* - \epsilon'$ , where  $\epsilon' = \epsilon + f_r(\mathbf{x}^*) - f_r(\mathbf{x}')$ . Similarly,  $\mathbf{x}^{*T} C \mathbf{y}^* \geq \mathbf{x}^{*T} C \mathbf{y}' - \epsilon''$ , where  $\epsilon'' = \epsilon + f_c(\mathbf{y}^*) - f_c(\mathbf{y}')$ . But  $\epsilon' = \epsilon'' = \epsilon$  when  $n \rightarrow \infty$ . Hence  $(\mathbf{x}^*, \mathbf{y}^*)$  is a  $\epsilon$ -NE for the bimatrix game  $(R, C)$ . This means that if there is an FPTAS for computing an  $\epsilon$ -equilibrium in a penalty game in  $\mathcal{P}'$  then there is an FPTAS for computing an  $\epsilon$ -NE in  $(R, C)$  which is a contradiction, unless  $P = PPA$ .  $\square$

## B Proof of Lemma 1

*Proof.* From Theorem 2 we know that for the chosen value of  $k$  there exists a  $k$ -uniform point  $\mathbf{x}'$  such that  $\|\mathbf{x}' - \mathbf{x}^*\|_p < \epsilon/\lambda$ . Since the function  $f(\mathbf{x})$  is  $\lambda_p$ -Lipschitz continuous, we get that  $|f(\mathbf{x}') - f(\mathbf{x}^*)| < \epsilon$ . In order to compute this point we have to exhaustively evaluate the function  $f$  in all  $k$ -uniform points and choose the point that it maximizes/minimizes its value. Since there are  $\binom{n+k-1}{k} = O(n^k)$  possible  $k$ -uniform points, the theorem follows.  $\square$

## C Proof of Lemma 2

*Proof.* Suppose that both events  $\phi_i$  and  $\bigcap_j \psi_j \in [M]$  hold. We will show that the event  $\pi_i$  must be true too. Let  $x_i$  be an arbitrary strategy, let  $\mathbf{x}_{-i}^*$  be a strategy profile for the rest of the players, and let  $\mathbf{x}_{-i}'$  be a sampled strategy profile from  $\mathbf{x}_{-i}^*$ . Since we assume that the events  $\psi_j$  is true for all  $j$  we get  $\|\mathbf{x}_{-i}' - \mathbf{x}_{-i}^*\|_p \leq \sum_{j \neq i} \|x_j' - x_j^*\|_p$  we get that

$$\begin{aligned} \|\mathbf{x}_{-i}' - \mathbf{x}_{-i}^*\|_p &\leq \sum_{j \neq i} \|x_j' - x_j^*\|_p \\ &\leq \sum_{j \neq i} \frac{\epsilon}{2M\lambda} \\ &\leq \frac{\epsilon}{2\lambda}. \end{aligned}$$

Furthermore, since by assumption the utility functions for the players are  $\lambda_p$ -Lipschitz continuous we have that

$$|T_i(x_i, \mathbf{x}_{-i}') - T_i(x_i, \mathbf{x}_{-i}^*)| \leq \frac{\epsilon}{2}.$$

This means that

$$\begin{aligned} T_i(x_i, \mathbf{x}'_{-i}) &\leq T_i(x_i, \mathbf{x}^*_{-i}) + \frac{\epsilon}{2} \\ &\leq T_i(x_i^*, \mathbf{x}^*_{-i}) + \frac{\epsilon}{2} \end{aligned} \quad (1)$$

since  $T_i(x_i^*, \mathbf{x}^*_{-i}) \geq T_i(x_i, \mathbf{x}^*_{-i})$  for all possible  $x_i$ ; the strategy profile  $(x_i^*, \mathbf{x}^*_{-i})$  is an equilibrium of the game. Furthermore, since by assumption the event  $\phi_i$  is true we get that

$$T_i(x_i^*, \mathbf{x}^*_{-i}) < T_i(x'_i, \mathbf{x}'_{-i}) + \frac{\epsilon}{2}. \quad (2)$$

Hence, if we combine the inequalities (1) and (2) we get that  $T_i(x_i, \mathbf{x}'_{-i}) < T_i(x'_i, \mathbf{x}'_{-i}) + \epsilon$  for all possible  $x_i$ . Thus, if the events  $\phi_i$  and  $\psi_j$  for every  $j \in [M]$  hold, then the event  $\pi_i$  holds too.  $\square$

## D Proof of Theorem 3

*Proof.* In order to prove the claim, it suffices to show that there is a strategy profile where every player plays a  $k$ -uniform strategy, for the chosen value of  $k$ , such that the events  $\pi_i$  hold for all  $i \in [M]$ . Since the utility functions in  $\mathfrak{L}$  are  $\lambda_p$ -Lipschitz continuous it holds that  $\bigcap_{i \in [n]} \psi_i \subseteq \bigcap_{i \in [n]} \phi_i$ . Furthermore, combining that with Lemma 2 we get that  $\bigcap_{i \in [n]} \psi_i \subseteq \bigcap_{i \in [n]} \pi_i$ . Thus, if the event  $\psi_i$  is true for every  $i \in [n]$ , then the event  $\bigcap_{i \in [n]} \pi_i$  is true as well.

From Theorem 2 we get that for each  $i \in [M]$  there is a  $\frac{16M^2\lambda^2 p\gamma^2}{\epsilon^2}$ -uniform point  $x'_i$  such that the event  $\psi_i$  occurs with positive probability. The claim follows.  $\square$

## E Proof of Lemma 4

*Proof.* We begin by introducing the following auxiliary events for all  $i \in [n]$

$$\psi_{ri} = \left\{ R_i \mathbf{y}' < R_i \mathbf{y}^* + \frac{\epsilon}{2} \right\}.$$

We prove how the events  $\psi_{ri}$  and the event  $\phi_r$  are related with the event  $\pi_r$ . Assume that the event  $\phi_r$  and the events  $\psi_{ri}$  for all  $i \in [n]$  are true. Let  $\mathbf{x}$  be any mixed strategy for the row player. Since by assumption  $R_i \mathbf{y}' < R_i \mathbf{y}^* + \frac{\epsilon}{2}$  and since  $\mathbf{x}$  is a probability distribution, it holds that  $\mathbf{x}^T R \mathbf{y}' < \mathbf{x}^T R \mathbf{y}^* + \frac{\epsilon}{2}$ . If we subtract  $f_r(\mathbf{x})$  from each side we get that  $\mathbf{x}^T R \mathbf{y}' - f_r(\mathbf{x}) < \mathbf{x}^T R \mathbf{y}^* - f_r(\mathbf{x}) + \frac{\epsilon}{2}$ . This means that  $T_r(\mathbf{x}, \mathbf{y}') < T_r(\mathbf{x}, \mathbf{y}^*) + \frac{\epsilon}{2}$  for all  $\mathbf{x}$ . But we know that  $T_r(\mathbf{x}, \mathbf{y}^*) \leq T_r(\mathbf{x}^*, \mathbf{y}^*)$  for all  $\mathbf{x} \in \Delta^n$ , since  $(\mathbf{x}^*, \mathbf{y}^*)$  is an equilibrium. Thus, we get that  $T_r(\mathbf{x}, \mathbf{y}') < T_r(\mathbf{x}^*, \mathbf{y}^*) + \frac{\epsilon}{2}$  for all possible  $\mathbf{x}$ . Furthermore, since the event  $\phi_r$  is true too, we get that  $T_r(\mathbf{x}, \mathbf{y}') < T_r(\mathbf{x}', \mathbf{y}') + \epsilon$ . Thus, if the events  $\phi_r$  and  $\psi_{ri}$  for all  $i \in [n]$  are true, then the event  $\pi_r$  must be true as well. Formally,  $\phi_r \bigcap_{i \in [n]} \psi_{ri} \subseteq \pi_r$ . Thus,  $Pr(\pi_r^c) \leq Pr(\phi_r^c) + \sum_i \psi_{ri}$ . Using the Hoeffding bound, we get that  $Pr(\psi_{ri}^c) \leq e^{-\frac{k\epsilon^2}{2}}$  for all  $i \in [n]$ . Our claim follows.  $\square$

## F Proof of Lemma 5

*Proof.* Since we assume that the penalty function  $f_r(\mathbf{x}')$  is  $\lambda_p$ -Lipschitz continuous the event  $\phi_{rb}$  can be replaced by the event  $\phi_{rb'} = \{\|\mathbf{x}' - \mathbf{x}^*\|_p < \epsilon/4\lambda\}$ . It is easy to see that  $\phi_{rb} \subseteq \phi_{rb'}$ . Then, using the proof of Theorem 2 from [3] we get that  $E[\|\mathbf{x}' - \mathbf{x}^*\|_p] \leq \frac{2\sqrt{p}}{\sqrt{k}}$ . Thus, using Markov's inequality we get that

$$\begin{aligned} Pr(\|\mathbf{x}' - \mathbf{x}^*\|_p \geq \frac{\epsilon}{4\lambda}) &\leq \frac{E[\|\mathbf{x}' - \mathbf{x}^*\|_p]}{\frac{\epsilon}{4\lambda}} \\ &\leq \frac{8\lambda\sqrt{p}}{\epsilon\sqrt{k}}. \end{aligned}$$

□

## G Proof of Theorem 6

*Proof.*

$$\begin{aligned} Pr(GOOD^c) &\leq Pr(\phi_r^c) + Pr(\pi_r^c) + Pr(\phi_c^c) + Pr(\pi_c^c) \\ &\leq 2(Pr(\phi_r^c) + Pr(\pi_r^c)) \\ &\leq 2(2Pr(\phi_r^c) + n \cdot e^{-\frac{k\epsilon^2}{2}}) \quad (\text{from Lemma 4}) \\ &\leq 2(2Pr(\phi_{ru}^c) + 2Pr(\phi_{rb'}^c) + n \cdot e^{-\frac{k\epsilon^2}{2}}) \\ &\leq 2(4e^{-\frac{k\epsilon^2}{8}} + \frac{8\lambda\sqrt{p}}{\epsilon\sqrt{k}} + n \cdot e^{-\frac{k\epsilon^2}{2}}) \quad (\text{from Lemma 5}) \\ &< 1 \quad \text{for the chosen value of } k. \end{aligned}$$

Hence,  $Pr(GOOD) > 0$  and our claim follows. □

## H Proof of Lemma 6

*Proof.* We will prove that  $|T_r(\hat{\mathbf{x}}, \mathbf{y}) - T_r(\mathbf{x}', \mathbf{y})| < \epsilon$  which implies our claim. Let  $\phi_1 = \{|\hat{\mathbf{x}}^T R\mathbf{y} - \mathbf{x}'^T R\mathbf{y}| \leq \epsilon/2\}$  and  $\phi_2 = \{|f_r(\hat{\mathbf{x}}) - f_r(\mathbf{x}')| < \epsilon/2\}$ . Notice that Lemma 5 does not use anywhere the fact that  $\mathbf{x}^*$  is an equilibrium strategy, thus it holds even if  $\mathbf{x}^*$  is replaced by  $\hat{\mathbf{x}}$ . Thus,  $Pr(\phi_2^c) \leq \frac{4\lambda\sqrt{p}}{\epsilon\sqrt{k}}$ . Furthermore, using the analysis from [16] again, we can prove that  $Pr(\phi_1^c) \leq 2e^{-\frac{k\epsilon^2}{4}}$  and using similar arguments as in the proof of Theorem 6 it can be easily proved that for the chosen of  $l$  it holds that  $Pr(\phi_1^c) + Pr(\phi_2^c) < 1$ , thus the events  $\phi_1$  and  $\phi_2$  occur with positive probability and our claim follows. □

## I Distance Biased Games

In this section, we focus on three particular classes of distance biased games, and we provide polynomial-time approximation algorithms for these games. We focus on the following three penalty functions:

- $L_1$  **penalty**:  $b_r(\mathbf{x}, \mathbf{p}) = \|\mathbf{x} - \mathbf{p}\|_1 = \sum_i |\mathbf{x}_i - \mathbf{p}_i|$ .
- $L_2^2$  **penalty**:  $b_r(\mathbf{x}, \mathbf{p}) = \|\mathbf{x} - \mathbf{p}\|_2^2 = \sum_i (\mathbf{x}_i - \mathbf{p}_i)^2$ .
- $L_\infty$  **penalty**:  $b_r(\mathbf{x}, \mathbf{p}) = \|\mathbf{x} - \mathbf{p}\|_\infty = \max_i |\mathbf{x}_i - \mathbf{p}_i|$ .

Our approach is to follow the well-known technique of [10] that finds a 0.5-NE in a bimatrix game. The algorithm that we will use for all three penalty functions is given below.

### Algorithm 4. The Base Algorithm

1. Compute a best response  $\mathbf{y}^*$  against  $\mathbf{p}$ .
2. Compute a best response  $\mathbf{x}$  against  $\mathbf{y}^*$ .
3. Set  $\mathbf{x}^* = \delta \cdot \mathbf{p} + (1 - \delta) \cdot \mathbf{x}$ , for some  $\delta \in [0, 1]$ .
4. Return the strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$ .

While this is a well-known technique for bimatrix games, note that it cannot immediately be applied to penalty games. This is because the algorithm requires us to compute two best response strategies, and while computing a best-response is trivial in bimatrix games, this is not the case for penalty games. Best responses for  $L_1$  and  $L_\infty$  penalties can be computed in polynomial-time via linear programming, and for  $L_2^2$  penalties, the ellipsoid algorithm can be applied. However, these methods do not provide strongly polynomial algorithms.

In this section, for each of the penalties, we develop a simple combinatorial algorithm for computing best response strategies for each of these penalties. Our algorithms are strongly polynomial. Then, we determine the quality of the approximation given by the base algorithm when our best response techniques are used. In what follows we make the common assumption that the payoffs of the underlying bimatrix game  $(R, C)$  are in  $[0, 1]$ .

### I.1 A 2/3-approximation algorithm for $L_1$ -biased games

We start by considering  $L_1$ -biased games. Suppose that we want to compute a best-response for the row player against a fixed strategy  $\mathbf{y}$  of the column player. We will show that best response strategies in  $L_1$ -biased games have a very particular form: if  $b$  is the best response strategy in the (unbiased) bimatrix game  $(R, C)$ , then the best-response places all of its probability on  $b$  *except* for a certain set of rows  $S$  where it is too costly to shift probability away from  $\mathbf{p}$ . The rows  $i \in S$  will be played with  $\mathbf{p}_i$  to avoid taking the penalty for deviating.

The characterisation for whether it is too expensive to shift away from  $\mathbf{p}$  is given by the following lemma.



**Lemma 7.** *Let  $j$  be a pure strategy, let  $k$  be a pure strategy with  $\mathbf{p}_k > 0$ , and let  $\mathbf{x}$  be a strategy with  $\mathbf{x}_k = \mathbf{p}_k$ . The utility for the row player increases when we shift probability from  $k$  to  $j$  if and only if  $R_j \mathbf{y} - R_k \mathbf{y} - 2d_r > 0$ .*

*Proof.* Suppose that we shift  $\delta$  probability from  $k$  to  $j$ , where  $\delta \in (0, \mathbf{p}_k]$ . Then the utility for the row player is equal to  $T_r(\mathbf{x}, \mathbf{y}) + \delta \cdot (R_j \mathbf{y} - R_k \mathbf{y} - 2d_r)$ , where the final term is the penalty for shifting away from  $k$ . Thus, the utility for the row player increases under this shift if and only if  $R_j \mathbf{y} - R_k \mathbf{y} - 2d_r > 0$ .  $\square$

Observe that, if we are able to shift probability away from a strategy  $k$ , then we should obviously shift it to a best response strategy for the (unbiased) bimatrix game, since this strategy maximizes the increase in our payoff. Hence, our characterisation of best response strategies is correct. This gives us the following simple algorithm for computing best responses.

**Algorithm 5.** Best Response Algorithm for  $L_1$  penalty

1. Set  $S = 0$ .
2. Compute a best response  $b$  against  $\mathbf{y}$  in the unbiased bimatrix game  $(R, C)$ .
3. For each index  $i \neq b$  in the range  $1 \leq i \leq n$ :
  - (a) If  $R_b \cdot \mathbf{y} - R_i \cdot \mathbf{y} - 2d_r \leq 0$ , then set  $\mathbf{x}_i = \mathbf{p}_i$  and  $S = S + \mathbf{p}_i$ .
  - (b) Otherwise set  $\mathbf{x}_i = 0$
4. Set  $\mathbf{x}_b = 1 - S$ .
5. Return  $\mathbf{x}$ .

Our characterisation has a number of consequences. Firstly, it can be seen that if  $d_r \geq 1/2$ , then there is no profitable shift of probability between any two pure strategies, since  $0 \leq R_i \mathbf{y} \leq 1$  for all  $i \in [n]$ . Thus, we get the following corollary.

**Corollary 1.** *If  $d_r \geq 1/2$ , then  $\mathbf{p}$  is a dominant strategy.*

Moreover, since we can compute a best response in polynomial time we get the next theorem.

**Theorem 9.** *In biased games with  $L_1$  penalty functions and  $\max\{d_r, d_c\} \geq 1/2$ , an equilibrium can be computed in polynomial time.*

Finally, using the characterization of best responses we can see that there is a connection between the equilibria of the distance biased game and the well supported Nash equilibria (WSNE) of the underlying bimatrix game.

**Theorem 10.** *Let  $\mathcal{B} = (R, C, \mathbf{b}_r(\mathbf{x}, \mathbf{p}), \mathbf{b}_c(\mathbf{y}, \mathbf{q}), d_r, d_c)$  be a distance biased game with  $L_1$  penalties and let  $d := \max\{d_r, d_c\}$ . Any equilibrium of  $\mathcal{B}$  is a 2d-WSNE for the bimatrix game  $(R, C)$ .*

*Proof.* Let  $(\mathbf{x}^*, \mathbf{y}^*)$  be an equilibrium for  $\mathcal{B}$ . From the best response Algorithm for  $L_1$  penalty games we can see that  $\mathbf{x}_i^* > 0$  if and only if  $R_b \cdot \mathbf{y}^* - R_i \cdot \mathbf{y}^* - 2d_r \leq 0$ ,

where  $b$  is a pure best response against  $\mathbf{y}^*$ . This means that for every  $i \in [n]$  with  $\mathbf{x}_i^* > 0$ , it holds that  $R_i \cdot \mathbf{y}^* \geq \max_{j \in [n]} R_j \cdot \mathbf{y}^* - 2d$ . Similarly, it holds that  $C_i^T \cdot \mathbf{x}^* \geq \max_{j \in [n]} C_j^T \cdot \mathbf{x}^* - 2d$  for all  $i \in [n]$  with  $\mathbf{y}_i^* > 0$ . This is the definition of a  $2d$ -WSNE for the bimatrix game  $(R, C)$ .  $\square$

**Approximation algorithm** We now analyse the approximation guarantee provided by the base algorithm for  $L_1$ -biased games. So, let  $(\mathbf{x}^*, \mathbf{y}^*)$  be the strategy profile the is returned by the base algorithm. Since we have already shown that exact Nash equilibria can be found in games with either  $d_c \geq 1/2$  or  $d_r \geq 1/2$ , we will assume that both  $d_c$  and  $d_r$  are less than  $1/2$ , since this is the only interesting case.

We start by considering the regret of the row player. The following lemma will be used in the analysis of all three of our approximation algorithms.

**Lemma 8.** *Under the strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  the regret for the row player is at most  $\delta$ .*

*Proof.* Notice that for all  $i \in [n]$  we have

$$|\delta \mathbf{p}_i + (1 - \delta) \mathbf{x}_i - \mathbf{p}_i| = (1 - \delta) |\mathbf{x}_i - \mathbf{p}_i|,$$

hence  $\|\mathbf{x}^* - \mathbf{p}\|_1 = (1 - \delta) \|\mathbf{x} - \mathbf{p}\|_1$  and  $\|\mathbf{x}^* - \mathbf{p}\|_\infty = (1 - \delta) \|\mathbf{x} - \mathbf{p}\|_\infty$ . Furthermore, notice that  $\sum_i ((1 - \delta) \mathbf{x}_i + \delta \mathbf{p}_i - \mathbf{p}_i)^2 = (1 - \delta)^2 \|\mathbf{x} - \mathbf{p}\|_2^2$ , thus  $\|\mathbf{x}^* - \mathbf{p}\|_2^2 \leq (1 - \delta) \|\mathbf{x} - \mathbf{p}\|_2^2$ . Hence the payoff for the row player it holds  $T_r(\mathbf{x}^*, \mathbf{y}^*) \geq \delta \cdot T_r(\mathbf{p}, \mathbf{y}^*) + (1 - \delta) \cdot T_r(\mathbf{x}, \mathbf{y}^*)$  and his regret under the strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is

$$\begin{aligned} \mathcal{R}^r(\mathbf{x}^*, \mathbf{y}^*) &= \max_{\tilde{\mathbf{x}}} T_r(\tilde{\mathbf{x}}, \mathbf{y}^*) - T_r(\mathbf{x}^*, \mathbf{y}^*) \\ &= T_r(\mathbf{x}, \mathbf{y}^*) - T_r(\mathbf{x}^*, \mathbf{y}^*) \quad (\text{since } \mathbf{x} \text{ is a best response against } \mathbf{y}^*) \\ &\leq \delta (T_r(\mathbf{x}, \mathbf{y}^*) - T_r(\mathbf{p}, \mathbf{y}^*)) \\ &\leq \delta \quad (\text{since } \max_{\mathbf{x}} T_r(\mathbf{x}, \mathbf{y}^*) \leq 1 \text{ and } T_r(\mathbf{p}, \mathbf{y}^*) \geq 0). \end{aligned}$$

$\square$

Next, we consider the regret of the column player. The following lemma will be used for both the  $L_1$  case and the  $L_\infty$  case. Observe that in the  $L_1$  case, the precondition of  $d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q}) \leq 1$  always holds, since we have  $\|\mathbf{y}^* - \mathbf{q}\|_1 \leq 2$ , thus  $d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q}) \leq 1$  since we are only interested in the case where  $d_c \leq 1/2$ .

**Lemma 9.** *If  $d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q}) \leq 1$ , then under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  the column player suffers at most  $2 - 2\delta$  regret.*

*Proof.* The regret of the column player under the strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is

$$\begin{aligned}
\mathcal{R}^c(\mathbf{x}^*, \mathbf{y}^*) &= \max_{\mathbf{y}} T_c(\mathbf{x}^*, \mathbf{y}) - T_c(\mathbf{x}^*, \mathbf{y}^*) \\
&= \max_{\mathbf{y}} \left\{ (1 - \delta)T_c(\mathbf{x}, \mathbf{y}) + \delta T_c(\mathbf{p}, \mathbf{y}) \right\} - (1 - \delta)T_c(\mathbf{x}, \mathbf{y}^*) - \delta T_c(\mathbf{p}, \mathbf{y}^*) \\
&\leq (1 - \delta) \left( \max_{\mathbf{y}} T_c(\mathbf{x}^*, \mathbf{y}) - T_c(\mathbf{x}, \mathbf{y}^*) \right) \quad (\text{since } \mathbf{y}^* \text{ is a best response against } \mathbf{p}) \\
&\leq (1 - \delta)(1 + d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q})) \quad (\text{since } \max_{\mathbf{x}} T_c(\mathbf{x}^*, \mathbf{y}) \leq 1) \\
&\leq (1 - \delta) \cdot 2 \quad (\text{since } d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q}) \leq 1).
\end{aligned}$$

□

To complete the analysis, we must select a value for  $\delta$  that equalises the two regrets. It can easily be verified that setting  $\delta = 2/3$  ensures that  $\delta = 2 - 2\delta$ , and so we have the following theorem.

**Theorem 11.** *In biased games with  $L_1$  penalties a  $2/3$ -equilibrium can be computed in polynomial time.*

## I.2 A 5/7-approximation algorithm for $L_2^2$ -biased games

We now turn our attention to biased games with an  $L_2^2$  penalty. Again, we start by giving a combinatorial algorithm for finding a best response. Throughout this section, we fix  $\mathbf{y}$  as a column player strategy, and we will show how to compute a best response for the row player.

Best responses in  $L_2^2$ -biased games can be found by solving a quadratic program, and actually this particular quadratic program can be solved via the ellipsoid algorithm [15]. We will give a simple combinatorial algorithm that uses the Karush-Kuhn-Tucker (KKT) conditions, and produces a closed formula for the solution. Hence, we will obtain a strongly polynomial time algorithm for finding best responses.

Our algorithm can be applied on  $L_2^2$  penalty functions and any value  $d_r$ , but for notation simplicity we describe our method for  $d_r = 1$ . Furthermore, we define  $\alpha_i := R_i \mathbf{y} + 2\mathbf{p}_i$  and we call  $\alpha_i$  as the payoff of pure strategy  $i$ . Then, the utility for the row player can be written as  $T_r(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \mathbf{x}_i \cdot \alpha_i - \sum_{i=1}^n \mathbf{x}_i^2 - \mathbf{p}^T \mathbf{p}$ . Notice that the term  $\mathbf{p}^T \mathbf{p}$  is a constant and it does not affect the solution of the best response; so we can exclude it from our computations. Thus, a best response for the row player against strategy  $\mathbf{y}$  is the solution of the following quadratic program

$$\begin{aligned}
&\text{maximize} && \sum_{i=1}^n \mathbf{x}_i \cdot \alpha_i - \sum_{i=1}^n \mathbf{x}_i^2 \\
&\text{subject to} && \sum_{i=1}^n \mathbf{x}_i = 1 \\
&&& \mathbf{x}_i \geq 0 \quad \text{for all } i \in [n].
\end{aligned}$$

The Lagrangian function for this problem is

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda, \mathbf{u}) = \sum_{i=1}^n \mathbf{x}_i \cdot \alpha_i - \sum_{i=1}^n \mathbf{x}_i^2 - \lambda \left( \sum_{i=1}^n \mathbf{x}_i - 1 \right) - \sum_{i=1}^n u_i \mathbf{x}_i$$

and the corresponding KKT conditions

$$\alpha_i - \lambda - 2\mathbf{x}_i - \mathbf{u}_i = 0 \quad \text{for all } i \in [n] \quad (3)$$

$$\sum_{i=1}^n \mathbf{x}_i = 1 \quad (4)$$

$$\mathbf{x}_i \geq 0 \quad \text{for all } i \in [n] \quad (5)$$

$$\mathbf{x}_i \cdot \mathbf{u}_i = 0 \quad \text{for all } i \in [n]. \quad (6)$$

Constraints (3)-(5) are the stationarity conditions and (6) are the complementarity slackness conditions. We say that strategy  $\mathbf{x}$  is a *feasible response* if it satisfies the KKT conditions. The obvious way to compute a best response is by exhaustively checking all  $2^n$  possible combinations for the complementarity conditions and choose the feasible response that maximizes the utility for a player. Next we prove how we can bypass the brute force technique and compute all best responses in polynomial time.

In what follows, without loss of generality, we assume that  $\alpha_1 \geq \dots \geq \alpha_n$ . That is, the pure strategies are ordered according to their payoffs. In the next lemma we prove that in every best response, if a player plays pure strategy  $l$  with positive probability, then he must play every pure strategy  $k$  with  $k < l$  with positive probability.

**Lemma 10.** *In every best response  $\mathbf{x}^*$  if  $\mathbf{x}_l^* > 0$  then  $\mathbf{x}_k^* > 0$  for all  $k < l$ .*

*Proof.* For the sake of contradiction suppose that there is a best response  $\mathbf{x}^*$  and a  $k < l$  such that  $\mathbf{x}_l^* > 0$  and  $\mathbf{x}_k^* = 0$ . Let us denote  $M = \sum_{i \neq \{l, k\}} \alpha_i \cdot \mathbf{x}_i^* - \sum_{i \neq \{l, k\}} \mathbf{x}_i^{*2}$ . Suppose now that we shift some probability, denoted by  $\delta$ , from pure strategy  $l$  to pure strategy  $k$ . Then his utility is  $T_r(\mathbf{x}^*, \mathbf{y}) = M + \alpha_l \cdot (\mathbf{x}_l^* - \delta) - (\mathbf{x}_l^* - \delta)^2 + \alpha_k \cdot \delta - \delta^2$ , which is maximized for  $\delta = \frac{\alpha_k - \alpha_l + 2\mathbf{x}_l^*}{4}$ . Notice that  $\delta > 0$  since  $\alpha_k \geq \alpha_l$  and  $\mathbf{x}_l^* > 0$ , thus the row player can increase his utility by assigning positive probability to pure strategy  $k$  which contradicts the fact that  $\mathbf{x}^*$  is a best response.  $\square$

Lemma 10 implies that there are only  $n$  possible supports that a best response can use. Indeed, we can exploit the KKT conditions to derive, for each candidate support, the exact probability that each pure strategy would be played. We derive the probability as a function of  $\alpha_i$ s and of the support size. Suppose that the KKT conditions produce a feasible response when we set the support to have size  $k$ . From condition (3) we get that  $\mathbf{x}_i = \frac{1}{2}(\alpha_i - \lambda)$  for all  $1 \leq i \leq k$  and zero else. But we know that  $\sum_{j=1}^k \mathbf{x}_j = 1$ . Thus we get that  $\sum_{j=1}^k \frac{1}{2}(\alpha_j - \lambda) = 1$  and if

we solve for  $\lambda$  get that  $\lambda = \frac{\sum_{j=1}^k \alpha_j - 2}{k}$ . This means that for all  $i \in [k]$  we get

$$\mathbf{x}_i = \frac{1}{2} \left( \alpha_i - \frac{\sum_{j=1}^k \alpha_j - 2}{k} \right). \quad (7)$$

So, our algorithm does the following. It loops through all  $n$  candidate supports for a best response. For each one, it uses Equation (7) to determine the probabilities, and then checks whether these satisfy the KKT conditions, and thus if this is a feasible response. If it is, then it is saved for in a list of feasible responses, otherwise it is discarded. After all  $n$  possibilities have been checked, the feasible response with the highest payoff is then returned.

**Algorithm 6.** Best Response Algorithm for  $L_2^2$  penalty

1. For  $i = 1 \dots n$ 
  - (a) Set  $\mathbf{x}_1 \geq \dots \geq \mathbf{x}_i > 0$  and  $\mathbf{x}_{i+1} = \dots = \mathbf{x}_n = 0$ .
  - (b) Check if there is a feasible response under these constraints.
  - (c) If so, add it to the list of feasible responses.
2. Among the feasible responses choose one with the highest utility.

**Approximation Algorithm** We now show that the base algorithm gives a 5/7-approximation when applied to  $L_2^2$ -penalty games. For the row player's regret, we can use Lemma 8 to show that the regret is bounded by  $\delta$ . However, for the column player's regret, things are more involved. We will show that the regret of the column player is at most  $2.5 - 2.5\delta$ . That analysis depends on the maximum entry of the base strategy  $\mathbf{q}$  and more specifically on whether  $\max_k \{\mathbf{q}_k\} \leq 1/2$  or not.

**Lemma 11.** *If  $\max_k \{\mathbf{q}_k\} \leq 1/2$ , then the regret the column player suffers under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is at most  $2.5 - 2.5\delta$ .*

*Proof.* Note that when  $\max_k \{\mathbf{q}_k\} \leq 1/2$ , then  $\mathbf{b}_c = \|\mathbf{y} - \mathbf{p}\|_2^2 \leq 1.5$  for all possible  $\mathbf{y}$ . Then, using the analysis from Lemma 9, along with the fact that  $d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q}) \leq 2$  for  $L_2^2$  penalties, and since by assumption  $d_c = 1$ , the claim follows.  $\square$

For the case where there is a  $k$  such that  $\mathbf{q}_k > 1/2$  a more involved analysis is needed. The first goal is to prove that under *any* strategy  $\mathbf{y}^*$  that is a best response against  $\mathbf{p}$  the pure strategy  $k$  is played with positive probability. In order to prove that, first it is proven that there is a feasible response against strategy  $\mathbf{p}$  where pure strategy  $k$  is played with positive probability. In what follows we denote  $\alpha_i := C_i^T \mathbf{p} + 2\mathbf{q}_i$ .

**Lemma 12.** *Let  $\mathbf{q}_k > 1/2$  for some  $k \in [n]$ . Then there is a feasible response where pure strategy  $k$  is played with positive probability.*

*Proof.* Note that  $\alpha_k > 1$  since by assumption  $\mathbf{q}_k > 1/2$ . Recall from Equation (7) that in a feasible response  $\mathbf{y}$  it holds that  $\mathbf{y}_i = \frac{1}{2} \left( \alpha_i - \frac{\sum_{j=1}^k \alpha_j - 2}{k} \right)$ .

In order to prove the claim it is sufficient to show that  $\mathbf{y}_k > 0$  when in the KKT conditions is set  $\mathbf{y}_i > 0$  for all  $i \in [k]$ . Or equivalently, to show that  $\alpha_k - \frac{\sum_{j=1}^k \alpha_j - 2}{k} = \frac{1}{k} ((k-1)\alpha_k + 2 - \sum_{j=1}^{k-1} \alpha_j) > 0$ . But,

$$\begin{aligned} (k-1)\alpha_k + 2 - \sum_{j=1}^{k-1} \alpha_j &> k+1 - \sum_{j=1}^{k-1} (C^T \mathbf{x} + 2\mathbf{q}_j) \quad (\text{since } \alpha_k > 1) \\ &\geq k+1 - (k-1) - \sum_{j=1}^{k-1} 2\mathbf{q}_j \\ &\geq 1 + \mathbf{q}_k \quad (\text{since } \mathbf{q} \in \Delta^n) \\ &> 0. \end{aligned}$$

The claim follows.  $\square$

Next it is proven that the utility of the column player is increasing when he adds pure strategies  $i$  in his support such that  $\alpha_i > 1$ .

**Lemma 13.** *Let  $\mathbf{y}^k$  and  $\mathbf{y}^{k+1}$  be two feasible responses with support size  $k$  and  $k+1$  respectively, where  $\alpha_{k+1} > 1$ . Then  $T_c(\mathbf{x}, \mathbf{y}^{k+1}) > T_c(\mathbf{x}, \mathbf{y}^k)$ .*

*Proof.* Let  $\mathbf{y}^k$  be a feasible response with support size  $k$  for the column player against strategy  $\mathbf{p}$  and let  $\lambda(k) := \frac{\sum_{j=1}^k \alpha_j - 2}{2k}$ . Then the utility of the column player when he plays  $\mathbf{y}^k$  can be written as

$$\begin{aligned} T_c(\mathbf{x}, \mathbf{y}^k) &= \sum_{i=1}^n \mathbf{y}_i^k \cdot \alpha_i - \sum_{i=1}^n (\mathbf{x}_i^k)^2 - \mathbf{q}^T \mathbf{q} \\ &= \sum_{i=1}^k \mathbf{y}_i^k (\alpha_i - \mathbf{y}_i^k) - \mathbf{q}^T \mathbf{q} \\ &= \sum_{i=1}^k \left( \frac{\alpha_i}{2} - \lambda(k) \right) \left( \frac{\alpha_i}{2} + \lambda(k) \right) - \mathbf{q}^T \mathbf{q} \\ &= \frac{1}{4} \sum_{i=1}^k \alpha_i^2 - k \cdot (\lambda(k))^2 - \mathbf{q}^T \mathbf{q}. \end{aligned}$$

The goal now is to prove that  $T_c(\mathbf{x}, \mathbf{y}^{k+1}) - T_c(\mathbf{x}, \mathbf{y}^k) > 0$ . By the previous analysis for  $T_c(\mathbf{x}, \mathbf{y}^k)$  and if  $A := \sum_{i=1}^k \alpha_i - 2$ , then

$$\begin{aligned}
T_c(\mathbf{x}, \mathbf{y}^{k+1}) - T_c(\mathbf{x}, \mathbf{y}^k) &= \frac{1}{4} \sum_{i=1}^{k+1} \alpha_i^2 - (k+1)(\lambda(k+1))^2 - \frac{1}{4} \sum_{i=1}^k \alpha_i^2 + k \cdot (\lambda(k))^2 \\
&= \frac{1}{4} \left( \alpha_{k+1}^2 + \frac{A^2}{k} - \frac{(A + \alpha_{k+1})^2}{k+1} \right) \\
&= \frac{1}{4} \left( \alpha_{k+1}^2 + \frac{1}{k+1} (A^2 - \alpha_{k+1}^2 - 2A\alpha_{k+1}) \right) \\
&= \frac{1}{4(k+1)} (k\alpha_{k+1}^2 + A^2 - 2A\alpha_{k+1}) \\
&> \frac{1}{4(k+1)} (k + A^2 - 2A) \quad (\text{since } 1 < \alpha_{k+1} \leq 2 \text{ and } A > k - 2) \\
&> \frac{1}{4(k+1)} (k^2 - 5k + 8) \quad (\text{since } A > k - 2) \\
&> 0.
\end{aligned}$$

□

Notice that  $\alpha_k \geq 2\mathbf{p}_k > 1$ . Thus, the utility of the feasible response that assigns positive probability to pure strategy  $k$  is strictly greater than the utility of any feasible responses that does not assign probability to  $k$ . Thus strategy  $k$  is always played in a best response. Hence, the next lemma follows.

**Lemma 14.** *If there is a  $k \in [n]$  such that  $\mathbf{q}_k > 1/2$ , then in every best response  $\mathbf{y}^*$  the pure strategy  $k$  is played with positive probability.*

Using now Lemma 14 we can provide a better bound for the regret the column player suffers, since in every best response  $\mathbf{y}^*$  the pure strategy  $k$  is played with positive probability.

**Lemma 15.** *Let  $\mathbf{y}^*$  be a best response when there is a pure strategy  $k$  with  $\mathbf{q}_k > 1/2$ . Then the regret for the column player under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is bounded by  $2 - 2\delta$ .*

*Proof.* Before we proceed with our analysis we assume without loss of generality that  $k = 1$ . Recall from the analysis for the Algorithm 1 that the regret for the column player is

$$\begin{aligned}
\mathcal{R}^c(\mathbf{x}^*, \mathbf{y}^*) &\leq (1 - \delta) \left( \max_{\tilde{\mathbf{y}} \in \Delta} \{\tilde{\mathbf{x}}^T C \tilde{\mathbf{y}}\} + 2\tilde{\mathbf{y}}^T \mathbf{q}_k - 2\mathbf{y}^{*T} \mathbf{q} + \mathbf{y}^{*T} \mathbf{y}^* \right) \\
&\leq (1 - \delta) (1 + 2\mathbf{q}_k - 2\mathbf{y}^{*T} \mathbf{q} + \mathbf{y}^{*T} \mathbf{y}^*).
\end{aligned} \tag{8}$$

We focus now on the term  $\mathbf{y}^{*T} \mathbf{y}^* - 2\mathbf{y}^{*T} \mathbf{q}$ . It can be proven <sup>3</sup> that  $\mathbf{y}^{*T} \mathbf{y}^* - 2\mathbf{y}^{*T} \mathbf{q} \leq 1 - 2\mathbf{q}_k$ . Thus, from (8) we get that  $\mathcal{R}^c(\mathbf{x}^*, \mathbf{y}^*) \leq 2 - 2\delta$ . □

<sup>3</sup> Appendix J

Recall now that the regret for the row player is bounded by  $\delta$ , so if we optimize with respect to  $\delta$  the regrets are equal for  $\delta = 2/3$ . Thus, the next theorem follows, since when there is a  $k$  with  $\mathbf{q}_k > 1/2$  the Algorithm 1 produces a  $2/3$ -equilibrium. Hence, combining this with Lemma 11 and Theorem 12 follows for  $\delta = 5/7$ .

**Theorem 12.** *In biased games with  $L_2^2$  penalties a  $5/7$ -equilibrium can be computed in polynomial time.*

### I.3 Inner product penalty games

We observe that we can also tackle the case where the penalty function is the inner product of the strategy played, i.e.  $\mathbf{p} = \mathbf{q} = \mathbf{0}$ . For these games, that we call inner product penalty games, we replace  $\mathbf{p}$  as the starting point of the base algorithm with the fully mixed strategy  $\mathbf{x}^n$ . Hence, for that case  $\mathbf{x}^* = \delta \cdot \mathbf{x}^n + (1 - \delta) \cdot \mathbf{x}$  for some  $\delta \in [0, 1]$ . Again, the regret the row player suffers under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is bounded by  $\delta$ .

**Lemma 16.** *When the penalty function is the inner product of the strategy played, then the regret for the row player under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is bounded by  $\delta$ .*

Furthermore, using similar analysis as in Lemma 9 it can be proven that the regret for the column player under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is bounded by  $(1 - \delta)(1 + d_c \cdot \mathbf{y}^{*T} \mathbf{y}^*)$ . For the column player we will distinguish between the cases where  $d_c \leq 1/2$  and  $d_c > 1/2$ . For the first case where  $d_c \leq 1/2$  it is easy to see that the algorithm produces a  $0.6$ -equilibrium. For the other case, when  $d_c > 1/2$ , first it is proven that there is no pure best response.

**Lemma 17.** *If the penalty for the column player is equal to  $\mathbf{y}^T \mathbf{y}$  and  $d_c > \frac{1}{2}$ , then there is no pure best response against any strategy of the row player.*

*Proof.* Let  $C_j$  to denote the payoff of the column player from his  $j$ -th pure strategy against some strategy  $\mathbf{x}$  played by the row player. For the sake of contradiction, assume that there is a pure best response for the column player where, without loss of generality, he plays only his first pure strategy. Suppose now that he shifts some probability to his second strategy, that is he plays the first pure strategy with probability  $x$  and the second pure strategy with probability  $1 - x$ . The utility for the column player under this mixed strategy is  $x \cdot C_1 + (1 - x) \cdot C_2 - d_c \cdot (x^2 + (1 - x)^2)$ , which is maximized for  $x = \frac{2d_c + C_1 - C_2}{4d_c}$ . Notice that  $x > 0$ , which means that the column player can deviate from the pure strategy and increase his utility. The claim follows.  $\square$

With Lemma 17 in hand, it can be proven that when  $d_c > 1/2$  the column player does not play any pure strategy with probability greater than  $3/4$ .

**Lemma 18.** *If  $d_c > 1/2$ , then in  $\mathbf{y}^*$  no pure strategy is played with probability greater than  $3/4$ .*



*Proof.* For the sake of contradiction suppose that there is a pure strategy  $i$  in  $\mathbf{y}^*$  that is played with probability greater than  $3/4$ . Furthermore, let  $k$  be the support size of  $\mathbf{y}^*$ . From Lemma 17, since  $d_c > 1/2$ , we know that there is no pure best response, thus  $k \geq 2$ . Then using Equation (7) we get that  $\frac{3}{4} < \frac{1}{2}(\alpha_i - \frac{\sum_{j=1}^k \alpha_j - 2}{k})$ . If we solve for  $\alpha_j$  we get that  $\alpha_i > \frac{3k-4}{2k-2} > 1$  which is a contradiction since when  $\mathbf{q} = \mathbf{0}$  it holds that  $\alpha_i = C_i^T \mathbf{x} \leq 1$ .  $\square$

A direct corollary from Lemma 18 is that  $\mathbf{y}^{*T} \mathbf{y}^* \leq 5/8$ . Hence, we can prove the following lemma.

**Lemma 19.** *Under strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  the regret for the column player is bounded by  $\frac{13}{8}(1 - \delta)$ .*

*Proof.* Firstly, note that  $T_c(\mathbf{x}^*, \mathbf{y}^*) = \delta \mathbf{x}^{nT} C \mathbf{y}^* + (1 - \delta) \mathbf{x}^T C \mathbf{y}^* - \mathbf{y}^{*T} \mathbf{y}^*$ . Moreover,  $\max_{\tilde{\mathbf{y}} \in \Delta} \{\mathbf{x}^{nT} C \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \tilde{\mathbf{y}}\} - T_c(\mathbf{x}^n, \mathbf{y}^*) = 0$ , since  $\mathbf{y}^*$  is a best response against  $\mathbf{x}^n$ . Finally, notice that  $0 \leq \mathbf{y}^T \mathbf{y} \leq 1$  for all  $\mathbf{y}$ . Thus, the regret for the column player is

$$\begin{aligned} \mathcal{R}^c(\mathbf{x}^*, \mathbf{y}^*) &= (1 - \delta) \left( \max_{\tilde{\mathbf{y}} \in \Delta} \{\mathbf{x}^T C \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \tilde{\mathbf{y}}\} - \mathbf{x}^T C \mathbf{y}^* + \mathbf{y}^{*T} \mathbf{y}^* \right) \\ &< (1 - \delta) \left( 1 + \frac{5}{8} \right). \end{aligned}$$

which matches the claimed result.  $\square$

If we combine Lemmas 16 and 19 and solve for  $\delta$  we can see that the regrets are equal for  $\delta = \frac{13}{21}$ . Thus, we get the following theorem for biased games where  $\mathbf{q} = \mathbf{0}$ .

**Theorem 13.** *The strategy profile  $(\mathbf{x}^*, \mathbf{y}^*)$  is a  $\frac{13}{21}$ -equilibrium for biased games with  $\mathbf{q} = \mathbf{0}$ .*

#### I.4 A 2/3-approximation for $L_\infty$ -biased games

Finally, we turn our attention to the  $L_\infty$  penalty. We start by giving a combinatorial algorithm for finding best responses. Similar to the best response Algorithm for the  $L_1$  penalty, the intuition is to start from the base strategy  $\mathbf{p}$  of the row player and shift probability from pure strategies with low payoff to pure strategies with higher payoff. This time though, the shifted probability will be distributed between the pure strategies with higher payoff.

Without loss of generality assume that  $R_1 \mathbf{y} \geq \dots \geq R_n \mathbf{y}$ , ie., that the strategies are ordered according to their payoff in the unbiased bimatrix game. The set of pure strategies of the row player can be partitioned into three disjoint sets according to the payoff they yield:

$$\begin{aligned} \mathcal{H} &:= \{i \in [n] : R_i \mathbf{y} = R_1 \mathbf{y}\} \\ \mathcal{M} &:= \{i \in ([n] \setminus \mathcal{H}) : R_1 \mathbf{y} - R_i \mathbf{y} - d_r < 0\} \\ \mathcal{L} &:= \{i \in [n] : R_1 \mathbf{y} - R_i \mathbf{y} - d_r > 0\}. \end{aligned}$$

Next we give an algorithm that computes a best response for  $L_\infty$  penalty.

**Algorithm 7.** Best Response Algorithm for  $L_\infty$  penalty

1. For all  $i \in \mathcal{L}$ , set  $x_i = 0$ .
2. If  $\mathcal{P} \leq |\mathcal{H}| \cdot p_{\max}$ , then set  $x_i = \mathbf{p}_i + \frac{\mathcal{P}}{|\mathcal{H}|}$  for all  $i \in \mathcal{H}$  and  $x_j = \mathbf{p}_j$  for  $j \in \mathcal{M}$ .
3. Else if  $\mathcal{P} < |\mathcal{H} \cup \mathcal{M}| \cdot p_{\max}$ , then
  - Set  $x_i = \mathbf{p}_i + p_{\max}$  for all  $i \in \mathcal{H}$ .
  - Set  $k = \lfloor \frac{\mathcal{P} - |\mathcal{H}| \cdot p_{\max}}{p_{\max}} \rfloor$ .
  - Set  $x_i = \mathbf{p}_i + p_{\max}$  for all  $i \leq |\mathcal{H}| + k$ .
  - Set  $x_{|\mathcal{H}|+k+1} = \mathbf{p}_{|\mathcal{H}|+k+1} + \mathcal{P} - (|\mathcal{H}| + k) \cdot p_{\max}$ .
  - Set  $x_j = \mathbf{p}_j$  for all  $|\mathcal{H}| + k + 2 \leq j \leq |\mathcal{H}| + |\mathcal{M}|$ .
4. Else set  $x_i = \mathbf{p}_i + \frac{\mathcal{P}}{|\mathcal{H} \cup \mathcal{M}|}$  for all  $i \in \mathcal{H} \cup \mathcal{M}$ .

Let  $p_{\max} := \max_{i \in \mathcal{L}} \mathbf{p}_i$  and let  $\mathcal{P} := \sum_{i \in \mathcal{L}} \mathbf{p}_i$ . Then for every best response the following lemma holds.

**Lemma 20.** *If  $\mathcal{L} \neq \emptyset$ , then for any best response  $\mathbf{x}$  of the row player against strategy  $\mathbf{y}$  it holds that  $\|\mathbf{x} - \mathbf{p}\|_\infty \geq p_{\max}$ . Else  $\mathbf{p}$  is the best response.*

*Proof.* Using similar arguments as in Lemma 7, it can be proven that if there are no pure strategies  $i$  and  $k$  such that  $R_k \mathbf{y} - R_i \mathbf{y} - d_r < 0$  then any shifting of probability decreases the utility of the row player. Thus, the best response of the player is  $\mathbf{p}$ . On the other hand, if there are strategies  $i$  and  $k$  such that  $R_k \mathbf{y} - R_i \mathbf{y} - d_r > 0$ , then the utility of the row player increase if all the probability from strategy  $i$  is shifted to pure strategy  $k$ . The set  $\mathcal{L}$  contains all these pure strategies. Let  $j \in \mathcal{L}$  be the pure strategy that defines  $p_{\max}$ . Then, all the  $p_{\max}$  probability can be shifted from  $j$  to the a pure strategy in  $\mathcal{H}$ , i.e. a pure strategy that yields the highest payoff, and strictly increase the utility of the player. Thus, the strategy  $j$  is played with zero probability and the claim follows.  $\square$

In what follows assume that  $\mathcal{L} \neq \emptyset$ , hence  $p_{\max} > 0$ . From Lemma 20 follows that there is a best response where the strategy with the highest payoff is played with probability  $\mathbf{p}_1 + p_{\max}$ . Hence, it can be shifted up to  $p_{\max}$  probability from pure strategies with lower payoff to each pure strategy with higher payoff, starting from the second pure strategy etc. After this shift of probabilities there will be a set of pure strategies that where each one is played with probability  $\mathbf{p}_i + p_{\max}$  and possibly one pure strategy  $j$  that is played with probability less or equal to  $\mathbf{p}_j$ . The question is whether more probability should be shifted from the low payoff strategies to strategies that yield higher payoff. The next lemma establishes that no pure strategy from  $\mathcal{L}$  is played with positive probability in any best response against  $\mathbf{y}$ .

**Lemma 21.** *In every best response against strategy  $\mathbf{y}$  all pure strategies  $i \in \mathcal{L}$  are played with zero probability.*

*Proof.* Let  $K$  denote the set of pure strategies that are played with positive probability after the first shifting of probabilities. Without loss of generality assume that each strategy  $i \in K$  is played with probability  $\mathbf{p}_i + p_{\max}$ . Then the utility of the player under this strategy is equal to  $U = \sum_{i \in K} (\mathbf{p}_i + p_{\max}) \cdot R_i \mathbf{y} - d_r \cdot p_{\max}$ . For the sake of contradiction, assume that there is one strategy  $j$  from  $\mathcal{L}$  that belongs to  $K$ . Suppose that probability  $\delta$  is shifted from the strategy  $j$  to the first pure strategy. Then the utility for the player is equal to  $U + \delta(R_1 \mathbf{y} - R_j \mathbf{y} - d_r) > U$ , since by definition of  $\mathcal{L}$   $R_1 \mathbf{y} - R_j \mathbf{y} - d_r > 0$ . Thus, the utility of the player is increasing if probability is shifted. Notice that the analysis holds even if the penalty is  $p_{\max} + \delta$  instead of  $p_{\max}$ , thus the claim follows.  $\square$

Thus, all the probability  $\mathcal{P}$  from strategies from  $\mathcal{L}$  should be shifted to strategies yield higher payoff. The question now is what is the optimal way to distribute that probability over the strategies with the higher payoff. Clearly, the same amount of probability should be shifted in all strategies in  $\mathcal{H}$  since it makes the penalty smaller. Furthermore, it is easy to see that the maximum amount of probability is shifted to strategies in  $\mathcal{H}$ . Next we prove that if  $\mathcal{P} \geq p_{\max} \cdot (|\mathcal{H}| + |\mathcal{M}|)$  then  $\mathcal{P}$  is uniformly distributed over the pure strategies in  $\mathcal{H} \cup \mathcal{M}$ .

*Proof.* If  $\mathcal{P} \geq p_{\max} \cdot (|\mathcal{H}| + |\mathcal{M}|)$  then there is a best response where the probability  $\mathcal{P}$  is uniformly distributed over the pure strategies in  $\mathcal{H} \cup \mathcal{M}$ .

*Proof.* Let  $|\mathcal{H}| + |\mathcal{M}| = k$  and  $S = \mathcal{P} - k \cdot p_{\max}$ . Let

$$U = \sum_{i \in \mathcal{H} \cup \mathcal{M}} (\mathbf{p}_i + p_{\max} + \frac{S}{k}) R_i \mathbf{y} - d_r (p_{\max} + \frac{S}{k})$$

be the utility when the probability  $S$  is distributed uniformly over all pure strategies in  $\mathcal{H} \cup \mathcal{M}$ . Furthermore, let  $U'$  be the utility when  $\delta > 0$  probability is shifted from a pure strategy  $j$  to the first pure strategy that yields the highest payoff. Then  $U' = U + \delta(R_1 \mathbf{y} - R_j \mathbf{y} - d_r)$ , but  $R_1 \mathbf{y} - R_j \mathbf{y} - d_r \leq 0$  since  $j \in \mathcal{H} \cup \mathcal{M}$ . The claim follows.  $\square$

Using the previous analysis the correctness of the algorithm follows.

Note that, using similar arguments as in Lemma 7 the next lemma can be proved.

**Lemma 22.** *If  $d_r \geq 1$ , then  $\mathbf{p}$  is a dominant strategy.*

Furthermore, the combination of Lemma 22 with the fact that best responses can be computed in polynomial time gives the next theorem.

**Theorem 14.** *In biased games with  $L_\infty$  penalty functions and  $\max\{d_r, d_c\} \geq 1$ , an equilibrium can be computed in polynomial time.*

Again we can see that there is a connection between the equilibria of the distance biased game and the well supported Nash equilibria (WSNE) of the underlying bimatrix game.

**Observation 1** Let  $\mathcal{B} = (R, C, \mathbf{b}_r(\mathbf{x}, \mathbf{p}), \mathbf{b}_c(\mathbf{y}, \mathbf{q}), d_r, d_c)$  be a distance biased game with  $L_\infty$  penalties and let  $d := \max\{d_r, d_c\}$ . Any equilibrium of  $\mathcal{B}$  is a  $d$ -WSNE for the bimatrix game  $(R, C)$ .

**Approximation algorithm** For the quality of approximation, we can reuse the results that we proved for the  $L_1$  penalty. Lemma 8 applies unchanged. For Lemma 9, we observe that  $d_c \cdot \mathbf{b}_c(\mathbf{y}^*, \mathbf{q}) \leq 1$  when the penalty  $\mathbf{b}_c(\mathbf{y}^*, \mathbf{q})$  is the  $L_\infty$  norm, since for this case it holds  $\|\mathbf{y}^* - \mathbf{q}\|_\infty \leq 1$  and it is assumed that  $d_c \leq 1$ . Thus, we have the following theorem.

**Theorem 15.** In biased games with  $L_\infty$  penalties a  $2/3$ -equilibrium can be computed in polynomial time.

## J Proof that $\mathbf{y}^{*T} \mathbf{y}^* - 2\mathbf{y}_k^* \mathbf{q}_k \leq 1 - 2\mathbf{q}_k$ .

*Proof.* Notice from (7) that for all  $i$  we get  $\mathbf{y}_i = \mathbf{y}_k + \frac{1}{2}(\alpha_i - \alpha_k)$ . Using that we can write the term  $\mathbf{y}^T \mathbf{y} = \sum_i \mathbf{y}_i^2$  as follows for a when  $\mathbf{y}$  has support size  $s$

$$\begin{aligned} \sum_{i=1}^s \mathbf{y}_i^2 &= \mathbf{y}_k^2 + \sum_{i \neq k} \mathbf{y}_i^2 \\ &= \mathbf{y}_k^2 + \sum_{i \neq k} \left( \mathbf{y}_k + \frac{1}{2}(\alpha_i - \alpha_k) \right)^2 \\ &= s\mathbf{y}_k^2 + \left( \sum_{i \neq k} (\alpha_i - \alpha_k) \right) \mathbf{y}_k + \frac{1}{4} \sum_{i \neq k} (\alpha_k - \alpha_i)^2. \end{aligned}$$

Then we can see that  $\mathbf{y}^{*T} \mathbf{y} - 2\mathbf{y}_k^{*T} \mathbf{q}_k$  is increasing as  $\mathbf{y}_k^*$  increases, since we know from Lemma 14 that  $\mathbf{y}_k^* > 0$ . This becomes clear if we take the partial derivative of  $\mathbf{y}^{*T} \mathbf{y} - 2\mathbf{y}_k^* \mathbf{q}_k$  with respect to  $\mathbf{y}_k^*$  which is equal to

$$\begin{aligned} 2s\mathbf{y}_k^* + \sum_{i \neq k} (\alpha_i - \alpha_k) - 2\mathbf{q}_k &= 2s\mathbf{y}_k^* + \sum_{i \neq k} 2(\mathbf{y}_i^* - \mathbf{y}_k^*) - 2\mathbf{q}_k \quad (\text{since } \mathbf{y}_i = \mathbf{y}_k + \frac{1}{2}(\alpha_i - \alpha_k)) \\ &= 2s\mathbf{y}_k^* + 2 \sum_{i \neq k} \mathbf{y}_i^* - 2(s-1)\mathbf{y}_k^* - 2\mathbf{q}_k \\ &= 2 \sum_{i=1}^s \mathbf{y}_i^* - 2\mathbf{q}_k \\ &= 2 - 2\mathbf{q}_k \\ &\geq 0 \quad (\text{since } \mathbf{y}_k^* > 0). \end{aligned}$$

Thus, the value of  $\mathbf{y}^{*T} \mathbf{y} - 2\mathbf{y}_k^* \mathbf{q}_k$  is maximized when  $\mathbf{y}_k^* = 1$  and our claim follows.  $\square$